

Package ‘metaITH’

Elise Merritt

2019-08-26

Package: metaITH

Title: A Tool for Analyzing Intratumor Heterogeneity at Multiple Levels

Version: 0.0.0.9000

Author: Elise Merritt, em575@scarletmail.rutgers.edu

Description: This package allows for analysis of intratumor heterogeneity at the DNA, RNA, and immunogenic levels.

Depends: R ($\geq 3.5.0$), ggplot2, reshape2, ape, phylobase, dplyr, matrixStats

License: MIT

Encoding: UTF-8

LazyData: true

RoxygenNote: 6.1.1

Suggests: knitr, rmarkdown

VignetteBuilder: knitr

R topics documented:

Signatures	2
z_score_calculations	2
anti_pd1_favor_signature	2
apoptosis_signature	3
drug_resistance_signature	4
emt_signature	5
hypoxia_signature	6
proliferation_signature	7
specified_geneset_signature	7
signature_analysis	8
Phylograms	10
dna_phylograms	10
immune_phylograms	11
rna_phylograms	12
snv_heatmaps	12
Multi-region Divergence and Diversity	13
multi_level_divergence_diversity	13
Utilities	14
metaITH_analysis	14
Index	18
References	19

Signatures

z_score_calculations

Description

Calculates z-scores for all genes in all tumor samples. This function must be run before doing any signature analyses.

Usage

```
z_score_calculations(rna_expression_matrix_file)
```

Arguments

- `rna_expression_matrix_file`: Tab delimited text file containing expression level of all genes in all tumor and normal samples.
- Gene names as rows with header “Gene” and samples as columns
- Headers of normal samples should end in “.nr” and headers of tumor samples should not end in “.nr”

Details

Takes in RNA expression matrix containing the expression level ($\log_2(\text{TPM}+1)$ or other measure) of all genes in all tumor and normal samples. Output is a matrix containing z-scores for all genes in the entire tumor samples and writes a tab-delimited text file. Z-score for each gene in each tumor sample is calculated using mean and standard deviation calculated from respective gene expression across all normal samples.

Value

This function returns one file:

1. A table of z-score values. First column contains all genes found in RNA expression matrix. Subsequent columns contain z-score of each gene for each sample.

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
z_score_calculations("RNA_matrix.txt")
read.table("z-scores_matrix.txt", sep="\t", header=T)[1:4,1:4]
```

```
##      Gene  brs1.t1r.  brs1.t2r.  brs1.t3r.
## 1 RN7SL1  0.5900522  0.58684943 -0.08013765
## 2 SFTPC   -0.3472531 -1.98528401  0.09348880
## 3 AKR1C1  -0.8676535 -0.28814552 -0.74276365
## 4 AKR1C2  -0.7285173 -0.06270846 -0.69599408
```

anti_pd1_favor_signature

Description

Calculates anti-PD1 favor score for each tumor sample using gene expression data [$\log_2(\text{TPM}+1)$] for 28 genes that comprise of anti-PD1 favor signature (Gibney et al., 2016). This score is a measure of responsiveness against anti-PD1 therapy. Higher score represents higher responsiveness.

Usage

```
anti_pd1_favor_signature(z_score_matrix_file)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. Running the function “`z_score_calculations`” will create this file.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of anti-PD1 favor scores for each tumor sample or region and a heatmap of the anti-PD1 favor scores. For each gene, Z- scores corresponding to its expression in tumor is calculated based on mean and standard deviation of its expression in the normal samples, and then anti-PD1 favor score for each tumor sample is defined as mean z-scores of 28 genes involved.

Value

This function returns two files:

1. A tab-delimited text file with anti-PD1 scores. First column is “Gene,” followed by one column for each tumor region or sample. Two rows, one for sample names, one called “anti-PD1 favor score” shows the anti-PD1 favor score for each tumor region or sample.
2. A .tiff image file of a heatmap

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
anti_pd1_favor_signature("z-scores_matrix.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## pdf
## 2
```

```
read.table("anti-PD1_favor_score.txt", sep="\t", header=T)[,1:4]
```

```
##           Gene brs1.t1r. brs1.t2r. brs1.t3r.
## 1 anti-PD1 favor score 0.473427 0.1254103 0.7149302
```

apoptosis_signature

Description

Apoptosis score is calculated for each tumor sample using gene expression data $[\log_2(\text{TPM}+1)]$ for 6 genes known to be associated with apoptosis. This score is a measure of apoptotic activity. Higher score represents more apoptosis.

Usage

```
apoptosis_signature(z_score_matrix_file)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. Running the function “`z_score_calculations`” will create this file.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of apoptosis scores for each tumor sample or region and a heatmap of the apoptotic scores. For each gene, Z- scores corresponding to its expression in tumor is calculated based on mean and standard deviation of its expression in the normal samples, and then apoptotic score for each tumor sample is defined as mean z-scores of 6 genes involved.

Value

This function returns two files:

1. A table of apoptosis scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Apoptosis score” shows the apoptosis score for each tumor region or sample.
2. A .tiff image file of a heatmap

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
apoptosis_signature("z-scores_matrix.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## pdf
## 2
```

```
read.table("Apoptosis_score.txt", sep="\t", header=T)[,1:4]
```

```
##           Gene brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Apoptosis score 0.5290366  1.116061 0.8882319
```

drug_resistance_signature

Description

Drug resistance score is calculated for each tumor sample using gene expression data $[\log_2(\text{TPM}+1)]$ for 23 genes that comprise of Drug resistance signature (Hou et al., 2012). This score is a measure of Pemetrexed drug resistance. Higher score represents higher resistance.

Usage

```
drug_resistance_signature(z_score_matrix_file)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. Running the function “`z_score_calculations`” will create this file.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of Drug resistance scores for each tumor sample or region and a heatmap of the Drug resistance scores. For each gene, Z- scores corresponding to its expression in tumor is calculated based on mean and standard deviation of its expression in the normal samples, and then Drug resistance score for each tumor sample is defined as mean z-scores of 23 genes involved.

Value

This function returns two files:

1. A tab delimited .txt file with Drug resistance scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Drug resistance score” shows the Drug resistance score for each tumor region or sample.
2. A .tiff image file of a heatmap

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
drug_resistance_signature("z-scores_matrix.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## pdf
## 2
```

```
read.table("Pemetrexed_resistance_score.txt", sep="\t", header=T)[,1:4]
```

```
##                               Gene brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Pemetrexed resistance score 0.4137076  1.293488  0.814833
```

emt_signature

Description

EMT score is calculated for each tumor sample using gene expression data $[\log_2(\text{TPM}+1)]$ for 60 genes that comprise of EMT signature (Ramaker et al., 2017). This gene signature includes 49 epithelial related genes (e-genes) and 11 mesenchymal related genes (m-genes). This score is a measure of epithelial/mesenchymal nature of the tumor. Higher score represents more mesenchymal like features.

Usage

```
emt_signature(z_score_matrix_file)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. Running the function “`z_score_calculations`” will create this file.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of EMT scores for each tumor sample and a heatmap of the EMT scores. For each of e-genes and m-genes, z- scores corresponding to its expression in tumor is calculated based on mean and standard deviation of its expression in the normal samples, and then EMT score for each tumor sample is defined as: $\text{EM score} = \text{av. Z-score of m-genes} - \text{av. z-score of e-genes}$

Value

This function returns three files:

1. A table of emt scores. First column is “Gene”, followed by one column for each tumor region or sample. Three rows, one for sample names, one called “Epithelial score” shows the epithelial score for each tumor region or sample, one called “Mesenchymal score” shows the mesenchymal score for each tumor region or sample.
2. A table of mesenchymal-epithelial scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “M-E score” shows the mesenchymal-epithelial score for each tumor region or sample.
3. A .tiff image file of a heatmap

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
emt_signature("z-scores_matrix.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## pdf
## 2
read.table("Mesenchymal-Epithelial_score.txt", sep="\t", header=T)[,1:4]

##           Gene brs1.t1r. brs1.t2r. brs1.t3r.
## 1 M-E score 0.5018744 -0.3082992 0.05229863
```

hypoxia__signature

Description

Hypoxia score is calculated for each tumor sample using gene expression data $[\log_2(\text{TPM}+1)]$ for 21 genes that comprise of Hypoxia signature (Buffa et al., 2010). This score is a measure of extent of hypoxia in tumors. Higher score represents more hypoxia.

Usage

```
hypoxia__signature(z_score_matrix_file)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. Running the function “`z_score_calculations`” will create this file.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of Hypoxia scores for each tumor sample or region and a heatmap of the Hypoxia scores. For each gene, Z- scores corresponding to its expression in tumor is calculated based on mean and standard deviation of its expression in the normal samples, and then Hypoxia score for each tumor sample is defined as mean z-scores of the genes involved.

Value

This function returns two files:

1. A table of hypoxia scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Hypoxia score” shows the hypoxia score for each tumor region or sample.
2. A .tiff image file of a heatmap

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
hypoxia__signature("z-scores_matrix.txt")

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## pdf
## 2
read.table("Hypoxia_score.txt", sep="\t", header=T)[,1:4]

##           Gene brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Hypoxia score 0.4360642  2.036502  0.570733
```

proliferation_signature

Description

Proliferation score is calculated for each tumor sample using gene expression data $[\log_2(\text{TPM}+1)]$ for 121 genes that comprise of Proliferation signature (Wilkerson et al., 2012). This score is a measure of proliferation potential. Higher score represents higher proliferation potential.

Usage

```
proliferation_signature(z_score_matrix_file)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. Running the function “`z_score_calculations`” will create this file.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of Proliferation scores for each tumor sample and a heatmap of the Proliferation scores. For each gene, Z- scores corresponding to its expression in tumor is calculated based on mean and standard deviation of its expression in the normal samples, and then Proliferation score for each tumor sample is defined as mean z-scores of the genes involved.

Value

This function returns two files:

1. A table of proliferation scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Proliferation score” shows the proliferation score for each tumor region or sample.
2. A .tiff image file of a heatmap

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
proliferation_signature("z-scores_matrix.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## pdf
## 2
```

```
read.table("Proliferation_score.txt", sep="\t", header=T)[,1:4]
```

```
##           Gene  brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Proliferation score 0.09291747 0.8304992 0.4802478
```

specified_geneset_signature

Description

z-score is calculated for each tumor sample using gene expression data $[\log_2(\text{TPM}+1)]$ for user specified number of genes (in the form of a gene list).

Usage

```
specified_geneset_signature(z_score_matrix_file, geneset)
```

Arguments

- `z_score_matrix_file`: A matrix of z-scores for all genes in all tumor samples and regions. Running the function “`z_score_calculations`” will create this file.
- `geneset`: A text file containing a list of genes supplied by the user, one gene per row, with the header “Gene”.

Details

Takes in a z-score matrix created by function `z_score_calculations`. Outputs a table of scores for each tumor sample and a heatmap of the scores. For each gene, z-scores corresponding to its expression in tumor is calculated based on mean and standard deviation of its expression in the normal samples, and then score for each tumor sample is defined as mean z-scores of the genes involved.

Value

This function returns two files:

1. A table of geneset scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Geneset” score” shows the geneset score for each tumor region or sample.
2. A .tiff image file of a heatmap

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
specified_geneset_signature("z-scores_matrix.txt", "proliferation_geneset.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## pdf
## 2
```

```
read.table("Geneset_score.txt", sep="\t", header=T)[,1:4]
```

```
##           Gene  brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Geneset score 0.09291747 0.8304992 0.4802478
```

signature_analysis

Description

Proliferation, hypoxia, EMT, apoptosis, anti-PD1 favor and drug resistance scores are calculated for each tumor sample using gene expression data $[\log_2(\text{TPM}+1)]$ for genes that comprise of respective signatures. This one function calculates all scores in one go.

Usage

```
signature_analysis(rna_expression_matrix_file)
```

Arguments

- `rna_expression_matrix_file`: Tab delimited file containing expression level of all genes in all tumor and normal samples.
- Gene names as rows with header “Gene” and samples as columns.
- Headers of normal samples should end in “.nr” and headers of tumor samples should not end in “.nr”.

Details

Takes in RNA expression matrix containing the expression level of all genes in all tumor samples and normals. Creates a z-score matrix and uses this matrix to find the hypoxia, proliferation, apoptosis, drug-resistance,

emt, and anti-PD1 favor scores for each tumor sample or region. Outputs a table and heatmap for each of these scores.

Value

This function returns thirteen files:

1. A table of hypoxia scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Hypoxia score” shows the hypoxia score for each tumor region or sample.
2. A .tiff image file visualizing the hypoxia scores as a heatmap
3. A table of proliferation scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Proliferation score” shows the proliferation score for each tumor region or sample.
4. A .tiff image file visualizing the proliferation scores as a heatmap
5. A table of apoptosis scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Apoptosis score” shows the apoptosis score for each tumor region or sample.
6. A .tiff image file visualizing the apoptosis scores as a heatmap
7. A table of drug-resistance scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Pemetrexed resistance score” shows the drug-resistance score for each tumor region or sample.
8. A .tiff image file visualizing the drug-resistance scores as a heatmap
9. A table of emt scores. First column is “Gene”, followed by one column for each tumor region or sample. Three rows, one for sample names, one called “Epithelial score” shows the epithelial score for each tumor region or sample, one called “Mesenchymal score” shows the mesenchymal score for each tumor region or sample.
10. A table of mesenchymal-epithelial scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “M-E score” shows the mesenchymal-epithelial score for each tumor region or sample.
11. A .tiff image file visualizing the mesenchymal-epithelial scores as a heatmap
12. A table of anti-PD1 favor scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “anti-PD1 favor score” shows the anti-PD1 favor score for each tumor region or sample.
13. A .tiff image file visualizing the anti-PD1 scores as a heatmap

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
signature_analysis("RNA_matrix.txt")
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
```

```
## Warning: Column `Gene` joining factors with different levels, coercing to
```

```
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector

## pdf
## 2

read.table("Proliferation_score.txt", sep="\t", header=T)[,1:4]

##           Gene  brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Proliferation score 0.09291747 0.8304992 0.4802478
```

Phylograms

dna_phylograms

Description

Calculates distance matrix from variant allele frequency data of different regions of a tumor and normal samples and generates a unrooted dendrogram from it using neighbor joining method. Apart from multi-region sequencing data, it could also be for any set of tumor or normal samples.

Usage

```
dna_phylograms(dna_phylo_list_file)
```

Arguments

- `dna_phylo_list_file`: A text file containing a list of sample files, one file name per line. Each file contains a matrix with variant allele frequency of all the variations in each tumor region and normal (see example datasets).

Details Takes in a text file containing list of file names. Each file contains a matrix with variant allele frequency of all the variations in each tumor region and normal. Creates a DNA distance matrix, containing the distance between all the tumor regions and normal, a DNA tree topology file, and an unrooted DNA phylogram. Distance is defined as the average difference of variant allele frequencies across all variations between any two regions (or samples).

Value

This function returns three files:

1. DNA distance matrix is a matrix of all samples v/s all samples. Each entry in the table corresponds to the distance between the row sample and column sample.
2. A text file containing information about the tree topology.
3. A .tiff image file of an unrooted dendrogram

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
dna_phylograms("DNA_phylo_list.txt")
read.table("DNA_distance_matrix_DNA_S1_frequency_matrix.txt", sep="\t", header=T)
```

##		R1	R2	R3	N
## R1		0.000000	4.814473	4.144641	15.52719
## R2		4.814473	0.000000	6.428683	16.06559
## R3		4.144641	6.428683	0.000000	13.74008
## N		15.527186	16.065588	13.740077	0.00000

immune__phylograms

Description

Calculates distance matrix from proportion of different immune cell types data (output obtained from CIBERSORT) of different regions of a tumor and normal samples and generates a unrooted dendrogram from it using neighbor joining method. Apart from multi-region sequencing data, it could also be for any set of tumor or normal samples.

Usage

```
immune__phylograms(immune__phylo__list__file)
```

Arguments

- immune__phylo__list__file: A text file containing a list of sample files, one file name per row. Each file contains a matrix of immune cell proportions of different immune cell types in each tumor region and normal (see example datasets).

Details

Takes in a text file containing list of file names. Each file contains a matrix with different immune cell type proportions in each tumor region and normal. Creates a immune distance matrix, containing the distance between all the tumor regions and normal, a immune tree topology file, and an unrooted immune phylogram. Distance is defined as the average difference of immune cell type proportions across all cell types between any two regions (or samples).

Value

This function returns three files:

1. Immune distance matrix is a matrix of all samples versus all samples. Each entry in the table corresponds to the distance between the row sample and column sample.
2. A text file containing information about the tree topology.
3. A .tiff image file of an unrooted dendrogram

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
immune__phylograms("Immune__phylo__list__txt")
read.table("Immune_distance_matrix_S1_Immune_CIBERSORT_matrix.txt", sep="\t", header=T)
```

##		R1	R2	R3	N
## R1		0.00000000	0.02679035	0.02137600	0.04413436
## R2		0.02679035	0.00000000	0.04783462	0.06250866
## R3		0.02137600	0.04783462	0.00000000	0.02289465
## N		0.04413436	0.06250866	0.02289465	0.00000000

rna_phylograms

Description

Calculates distance matrix from gene expression data ($\log_2(\text{TPM}+1)$) of different regions of a tumor and normal samples and generates a unrooted dendrogram from it using neighbor joining method. Apart from multi-region sequencing data, it could also be for any set of tumor or normal samples.

Usage

```
rna_phylograms(rna_phylo_list_file)
```

Arguments

- `rna_phylo_list_file`: A text file containing a list of sample files, one file name per row. Each file contains a matrix with gene expression data ($\log_2(\text{TPM}+1)$) of all the genes in each tumor region and normal (see example datasets).

Details

Takes in a text file containing list of file names. Each file contains a matrix with gene expression data of all genes in each tumor region and normal. Creates a RNA distance matrix, containing the distance between all the tumor regions and normal, a RNA tree topology file, and an unrooted RNA phylogram. Distance is defined as the average difference of gene expression values across all genes between any two regions (or samples).

Value

This function returns three files:

1. RNA distance matrix is a matrix of all samples versus all samples. Each entry in the table corresponds to the distance between the row sample and column sample.
2. A text file containing information about the tree topology.
3. A .tiff image file of an unrooted dendrogram

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
rna_phylograms("RNA_phylo_list.txt")
read.table("RNA_distance_matrix_S1_RNA_expresion_matrix.txt", sep="\t", header=T)
```

```
##           R1           R2           R3           N
## R1 0.0000000 0.6834438 0.409917 0.6300742
## R2 0.6834438 0.0000000 0.756774 1.2733578
## R3 0.4099170 0.7567740 0.000000 0.6876690
## N  0.6300742 1.2733578 0.687669 0.0000000
```

snv_heatmaps

Description

Generates a heatmap of variant allele frequencies of somatic variations in different regions of one tumor.

Usage `snv_heatmaps(dna_phylo_list_file)`

Arguments

- `dna_phylo_list_file`: A text file containing a list of sample files, one file name per row. Each file contains a matrix with variant allele frequency of all the variations in each tumor region and normal (see example datasets).

Details

Takes in a text file containing list of file names. Each file contains a matrix with variant allele frequency of all the variations in each tumor region and normal. Creates a heatmap of the variant allele frequencies.

Value

This function returns one file:

1. A .tiff image file of a heatmap

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
snv_heatmaps("DNA_phylo_list.txt")
```

Multi-region Divergence and Diversity

`multi_level_divergence_diversity`

Description

Provides an estimate of the ratio of intra-patient regional diversity to tumor-benign tissue divergence.

Usage

```
multi_level_divergence_diversity(sample_names_file)
```

Arguments

- `sample_names_file`: A text file containing list of sample names, one file name per line

Details Takes in text file containing list of samples, one sample per line, and distance matrix files generated by `dna_phylograms`, `rna_phylograms`, and `immune_phylograms` methods. Outputs a png file containing 6 plots, 3 diversity/divergence plots and 3 diversity plots. NOTE: Requires output of functions `dna_phylograms`, `rna_phylograms`, and `immune_phylograms` to be in same folder

Value

This function returns one file:

1. A .png image file

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
multi_level_divergence_diversity("sample_names.txt")
```

```
##
## Phylogenetic tree with 4 tips and 2 internal nodes.
##
## Tip labels:
## [1] "R1" "R2" "R3" "N"
##
## Unrooted; includes branch lengths.
##
## Phylogenetic tree with 4 tips and 2 internal nodes.
```

```
##
## Tip labels:
## [1] "R1" "R2" "R3" "N"
##
## Unrooted; includes branch lengths.
##
## Phylogenetic tree with 4 tips and 2 internal nodes.
##
## Tip labels:
## [1] "R1" "R2" "R3" "N"
##
## Unrooted; includes branch lengths.

## pdf
## 2
```

Utilities

metaITH_analysis

Description

Runs a comprehensive intra-tumor heterogeneity analyses including dendrogram analysis (using DNA, RNA, and immune expression data), signature analysis (using RNA expression data), and multi-region diversity and divergence analysis (using DNA, RNA, and immune expression data) in one go.

Usage

```
metaITH_analysis(dna_phylo_list, rna_phylo_list, immune_phylo_list, multi_region_sample_names,
rna_expression_matrix)
```

Arguments

- `dna_phylo_list`: A text file containing a list of sample files, one file name per line. Each file contains a matrix with variant allele frequency of all the variations in each tumor region and normal.
- `rna_phylo_list`: A text file containing a list of sample files, one file name per line. Each file contains a matrix with expression values (preferably $\log_2(\text{TPM}+1)$) of all the genes in each tumor region and normal.
- `immune_phylo_list`: A text file containing a list of sample files, one file name per line. Each file contains a matrix with proportion of immune cells (inferred by CIBERSORT) of all immune cell types in each tumor region and normal.
- `multi_region_sample_names`: A text file containing list of sample names for multi-region diversity and divergence analysis, one file name per line.
- `rna_expression_matrix`: File containing expression level of all genes in all tumor samples. Header of first column should be “Gene” and each normal file should end in “.nr”

Details

Analyzes several sample files by running a phylogram analysis (using DNA, RNA, and immune expression data), signature analysis (using RNA expression data), and multi-region diversity and divergence analysis (using DNA, RNA, and immune expression data).

Value

This function returns twenty-three files:

1. DNA distance matrix is a matrix of all samples versus all samples. Each entry in the table corresponds to the distance between the row sample and column sample.
2. A text file containing information about the DNA tree topology
3. A .tiff image file of an unrooted DNA dendrogram
4. RNA distance matrix is a matrix of all samples versus all samples. Each entry in the table corresponds to the distance between the row sample and column sample.
5. A text file containing information about the RNA tree topology
6. A .tiff image file of an unrooted RNA dendrogram
7. Immune distance matrix is a matrix of all samples versus all samples. Each entry in the table corresponds to the distance between the row sample and column sample.
8. A text file containing information about the immune tree topology
9. A .tiff image file showing an unrooted immune dendrogram
10. A .png image file containing 6 plots, 3 diversity/divergence plots and 3 diversity plots
11. A tab delimited .txt file with hypoxia scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Hypoxia score” shows the hypoxia score for each tumor region or sample.
12. A .tiff image file visualizing the hypoxia scores as a heatmap
13. A tab delimited .txt file with proliferation scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Proliferation score” shows the proliferation score for each tumor region or sample.
14. A .tiff image file visualizing the proliferation scores as a heatmap
15. A tab delimited .txt file with apoptosis scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Apoptosis score” shows the apoptosis score for each tumor region or sample.
16. A .tiff image file visualizing the apoptosis scores as a heatmap
17. A tab delimited .txt file with drug-resistance scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “Pemetrexed resistance score” shows the drug-resistance score for each tumor region or sample.
18. A .tiff formatted image file visualizing the drug-resistance scores as a heatmap
19. A tab delimited .txt file with epithelial and mesenchymal scores. First column is “Gene”, followed by one column for each tumor region or sample. Three rows, one for sample names, one called “Epithelial score” shows the epithelial score for each tumor region or sample, one called “Mesenchymal score” shows the mesenchymal score for each tumor region or sample.
20. A tab delimited .txt file with EMT scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “M-E score” shows the mesenchymal-epithelial score for each tumor region or sample.
21. A .tiff image file visualizing the EMT scores as a heatmap
22. A tab delimited text file with anti-PD1 favor scores. First column is “Gene”, followed by one column for each tumor region or sample. Two rows, one for sample names, one called “anti-PD1 favor score” shows the anti-PD1 favor score for each tumor region or sample.

23. A .tiff image file visualizing the anti-PD1 scores as a heatmap

Example

```
setwd("~/proliferative_index_files/newid_PC_analysis/metaITH/R")
metaITH_analysis("DNA_phylo_list.txt", "RNA_phylo_list.txt", "Immune_phylo_list.txt", "sample_names.txt")

##
## Phylogenetic tree with 4 tips and 2 internal nodes.
##
## Tip labels:
## [1] "R1" "R2" "R3" "N"
##
## Unrooted; includes branch lengths.
##
## Phylogenetic tree with 4 tips and 2 internal nodes.
##
## Tip labels:
## [1] "R1" "R2" "R3" "N"
##
## Unrooted; includes branch lengths.
##
## Phylogenetic tree with 4 tips and 2 internal nodes.
##
## Tip labels:
## [1] "R1" "R2" "R3" "N"
##
## Unrooted; includes branch lengths.
##
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
##
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
##
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
##
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
##
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
##
## Warning: Column `Gene` joining factors with different levels, coercing to
## character vector
##
## pdf
##      2
read.table("DNA_distance_matrix_DNA_S1_frequency_matrix.txt", sep="\t", header=T)

##           R1           R2           R3           N
## R1  0.000000  4.814473  4.144641 15.52719
```



```
## R2  4.814473  0.000000  6.428683 16.06559
## R3  4.144641  6.428683  0.000000 13.74008
## N   15.527186 16.065588 13.740077  0.00000
```

```
read.table("Proliferation_score.txt", sep="\t", header=T)[,1:4]
```

```
##           Gene  brs1.t1r. brs1.t2r. brs1.t3r.
## 1 Proliferation score 0.09291747 0.8304992 0.4802478
```

Index

Phylograms

- dna_phylograms, 10
- immune_phylograms, 10
- rna_phylograms, 11
- snv_heatmaps, 12

Signatures

- anti_pd1_favor_signature, 2
- apoptosis_signature, 3
- drug_resistance_signature, 4
- emt_signature, 5
- hypoxia_signature, 6
- proliferation_signature, 6
- signature_analysis, 8
- specified_geneset_analysis, 7
- z_score_calculations, 2

Multi-Level Divergence and Diversity

- multi_level_divergence_diversity, 13

Utilities

- metaITH_analysis, 14

References

1. Gibney, G.T., Weiner, L.M., and Atkins, M.B. (2016). Predictive biomarkers for checkpoint inhibitor-based immunotherapy. *Lancet. Oncol.* 17, e542–e551.
2. Hou, J., Lambers, M., den Hamer, B., den Bakker, M.A., Hoogsteden, H.C., Grosveld, F., Hegmans, J., Aerts, J., and Philipsen, S. (2012). Expression Profiling-Based Subtyping Identifies Novel Non-small Cell Lung Cancer Subgroups and Implicates Putative Resistance to Pemetrexed Therapy. *J. Thorac. Oncol.* 7, 105–114
3. Ramaker, R.C., Lasseigne, B.N., Hardigan, A.A., Palacio, L., Gunther, D.S., Myers, R.M., and Cooper, S.J. (2017). RNA sequencing-based cell proliferation analysis across 19 cancers identifies a subset of proliferation-informative cancers with a common survival signature. *Oncotarget* 8, 38668–38681.
4. Wilkerson, M.D., Yin, X., Hoadley, K.A., Liu, Y., Hayward, M.C., Cabanski, C.R., Muldrew, K., Miller, C.R., Randell, S.H., Socinski, M.A., et al. (2010). Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. *Clin. Cancer Res.* 16, 4864–4875.
5. Buffa, F. M., Harris, A. L., West, C. M., & Miller, C. J. (2010). Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. *British journal of cancer*, 102(2), 428–435.